

DIALIGN 2.1

m, Kornelie Frech,
ress Developed by Burkhard Morgenstern, Said Abdeddai
Klaus Hahn, Thomas Werner, Jens Stoye, Andreas D

e-mail: burkhard.morgenstern@rp-rorer.co.uk

cite: Published research assisted by DIALIGN 2 should

ent B. Morgenstern (1999),
"DIALIGN 2: improvement of the segment-to-segment
approach to multiple sequence alignment."
Bioinformatics 15, 203 - 210.

Options:

=====

- 1) nucleic acid sequences aligned
- 2) no translation of nucleotide diagonals into peptides
- 3) 5 "*" characters for regions of maximum similarity

Aligned sequences:

=====

length:

=====

- | | | |
|----|------------|------|
| 1) | EP | 2541 |
| 2) | 09/471,669 | 1503 |

Average sequence length: 2022.000

Please note that only upper-case letters are considered to be aligned.

For more information, have a look at the user guide

[http://bibiserv.techfak.uni-bielefeld.de/dialign/user_gu
ide2.html](http://bibiserv.techfak.uni-bielefeld.de/dialign/user_guide2.html)

Alignment (DIALIGN format);
=====

EP 1 ATGGCCCAAG CCCTGCCCTG GCTCCTGCTG TGGATGGGC
G CGGGAGTGCT 09/471,669 1 ATGGCCCAAG CCCTGCCCTG GCTCCTGCTG TGGATGGGC
G CGGGAGTGCT

* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****
* *****

EP 51 GCCTGCCAC GGCACCCAGC ACGGCATCCG GCTGCCCT
G CGCAGCGGCC 09/471,669 51 GCCTGCCAC GGCACCCAGC ACGGCATCCG GCTGCCCT
G CGCAGCGGCC

* *****
* *****
* *****
* *****
* *****
* *****

EP 101
C CGACGAAGAG
09/471,669 101
C CGACGAAGAG

TGGGGGGCGC CCCCCCTGGGG CTGCGGCTGC CCCGGGAGA
TGGGGGGCGC CCCCCCTGGGG CTGCGGCTGC CCCGGGAGA

* *****
* *****
* *****
* *****
* *****

EP 151
G TGGACAACCT
09/471,669 151
G TGGACAACCT

CCCGAGGGAGC CCGGCCGGAG GGGCAGCTTT GTGGAGATG
CCCGAGGGAGC CCGGCCGGAG GGGCAGCTTT GTGGAGATG

* *****
* *****
* *****
* *****
* *****

EP 201
C GTGGGCAGCC
09/471,669 201
C GTGGGCAGCC

GAGGGGCAAG TCAGGGCAGG GCTACTACGT GGAGATGAC
GAGGGGCAAG TCAGGGCAGG GCTACTACGT GGAGATGAC

* *****

* *****
* *****
*
*

EP 251
G TAACTTGCA
09/471,669 251
G TAACTTGCA

CCCCGCAGAC GCTAACATC CTGGTGGATA CAGGCAGCA
CCCCGCAGAC GCTAACATC CTGGTGGATA CAGGCAGCA

* *****
* *****
* *****
* *****
* *****

** *****
** *****

EP 301
C AGAGGCAGCT
09/471,669 301
C AGAGGCAGCT

GTGGGTGCTG CCCCCCACCC CTTCCTGCAT CGCTACTAC
GTGGGTGCTG CCCCCCACCC CTTCCTGCAT CGCTACTAC

* *****
* *****
* *****
* *****
* *****

EP 351
 G CCCTACACCC
 09/471,669 351
 G CCCTACACCC

GTCCAGCACA TACCGGGACC TCCGGAAGGG TGTGTATG^a
 GTCCAGCACA TACCGGGACC TCCGGAAGGG TGTGTATG^b

* *****
 * *****
 * *****
 * *****
 * *****
 * *****

EP 401
 G CATCCCCAT
 09/471,669 401
 G CATCCCCAT

AGGGCAAGTG GGAAGGGGAG CTGGGCACCG ACCTGGTAA
 AGGGCAAGTG GGAAGGGGAG CTGGGCACCG ACCTGGTAA

* *****
 * *****
 * *****
 * *****
 * *****

EP 451
 A CTGAATCAGA
 09/471,669 451
 A CTGAATCAGA

GGCCCCAACG TCACTGTGCG TGCCAACATT GCTGCCATC
 GGCCCCAACG TCACTGTGCG TGCCAACATT GCTGCCATC

* *****
 * *****

* *****
* *****
* *****

EP 501
G GGGCTGGCCT
09/471,669 501
G GGGCTGGCCT

CAAGTTCTTC ATCAACGGCT CCAACTGGGA AGGCATCCT
CAAGTTCTTC ATCAACGGCT CCAACTGGGA AGGCATCCT

* *****
* *****
* *****
* *****
* *****

EP 551
T CTTTGACTCT
09/471,669 551
T CTTTGACTCT

ATGCTGAGAT TGCCAGGCCT GACGACTCCC TGGAGCCTT
ATGCTGAGAT TGCCAGGCCT GACGACTCCC TGGAGCCTT

* *****
* *****
* *****
* *****
* *****

EP 601
C AGCTTTGTGG

CTGGTAAAGC AGACCCACGT TCCCAACCTC TTCTCCCTG

09/471,669 601
C AGCTTTGTGG

CTGGTAAAGC AGACCCACGT TCCCAACCTC TTCTCCCTG

* *****
* *****
* *****
* *****
* *****
* *****

EP 651
T GTCGGAGGGA
09/471,669 651
T GTCGGAGGGA

TGCTGGCTTC CCCCTCAACC AGTCTGAAGT GCTGGCCTC
TGCTGGCTTC CCCCTCAACC AGTCTGAAGT GCTGGCCTC

* *****
* *****
* *****
* *****
* *****

EP 701
G CAGTCTCTGG
09/471,669 701
G CAGTCTCTGG

GCATGATCAT TGGAGGTATC GACCACTCGC TGTACACAG
GCATGATCAT TGGAGGTATC GACCACTCGC TGTACACAG

* *****
* *****
* *****

* *****
* *****

EP 751
A TTGTGCGGGT
09/471,669 751
A TTGTGCGGGT

TATACACCCA TCCGGCGGGGA GTGGTATTAT GAGGTGATC
TATACACCCA TCCGGCGGGGA GTGGTATTAT GAGGTGATC

* *****
* *****
* *****
* *****
* *****
* *****

EP 801
G TACAACCTATG
09/471,669 801
G TACAACCTATG

GGAGATCAAT GGACAGGATC TGAAAATGGA CTGCAAGGA
GGAGATCAAT GGACAGGATC TGAAAATGGA CTGCAAGGA

* *****
* *****
* *****
* *****
* *****

EP 851
T GCCCAAGAAA
09/471,669 851
T GCCCAAGAAA

ACAAGAGCAT TGTGGACAGT GGCACCACCA ACCTTCGTT
ACAAGAGCAT TGTGGACAGT GGCACCACCA ACCTTCGTT

* *****
* *****
* *****
* *****
* *****
* *****

EP 901
T CCACGGAGAA
09/471,669 901
T CCACGGAGAA

GTGTTGAAG CTGCAGTCAA ATCCATCAAG GCAGCCTCC
GTGTTGAAG CTGCAGTCAA ATCCATCAAG GCAGCCTCC

* *****
* *****
* *****
* *****
* *****
* *****

EP 951
C TGGCAAGCAG
09/471,669 951
C TGGCAAGCAG

GTTCCCTGAT GGTTCTGGC TAGGAGAGCA GCTGGTGTG
GTTCCCTGAT GGTTCTGGC TAGGAGAGCA GCTGGTGTG

* *****
* *****
* *****
* *****

EP 1001
A CCTAATGGGT
09/471,669 1001
A CCTAATGGGT

GCACCA~~CCCC~~ TTGGAACATT TTCCCAGTCA TCTCACTCT
GCACCA~~CCCC~~ TTGGAACATT TTCCCAGTCA TCTCACTCT

EP 1051
C AGCAATAACCT
09/471,669 1051
C AGCAATAACCT

GAGGTTACCA ACCAGTCCTT CCGCATCACC ATCCTTCCG
GAGGTTACCA ACCAGTCCTT CCGCATCACC ATCCTTCCG

EP 1101
T TACAAGTTG
09/471,669 1101
T TACAAGTTG

GCGGCCAGTG GAAGATGTGG CCACGTCCCA AGACGACTG
GCGGCCAGTG GAAGATGTGG CCACGTCCCA AGACGACTG

* *****
* *****
* *****
* *****
* *****
* *****

EP 1151
T TATCATGGAG
09/471,669 1151
T TATCATGGAG

CCATCTCACA GTCATCCACG GGCACGTGTTA TGGGAGCTG
CCATCTCACA GTCATCCACG GGCACGTGTTA TGGGAGCTG

* *****
* *****
* *****
* *****
* *****
* *****

EP 1201
G GCTTGCTGT
09/471,669 1201
G GCTTGCTGT

GGCTTCTACG TTGTCTTGTA TCAGGGCCCGA AAACGAATT
GGCTTCTACG TTGTCTTGTA TCAGGGCCCGA AAACGAATT

EP 1251 CAGCGCTTGC CATGTGCACG ATGAGTTCA GACGGCAGC
G GTGGAAGGCC
09/471,669 1251 CAGCGCTTGC CATGTGCACG ATGAGTTCA GACGGCAGC
G GTGGAAGGCC

* *****
* *****
* *****
* *****
* *****
* *****
* *****

EP 1301 CTTTTGTCAC CTTGGACATG GAAGACTGTG GCTACAACA
T TCCACAGACA
09/471,669 1301 CTTTTGTCAC CTTGGACATG GAAGACTGTG GCTACAACA
T TCCACAGACA

* *****
* *****
* *****
* *****
* *****
* *****
* *****

EP 1351 GATGAGTCAA CCCTCATGAC CATAGCCTAT GTCATGGCT
G CCATCTGCGC
09/471,669 1351 GATGAGTCAA CCCTCATGAC CATAGCCTAT GTCATGGCT
G CCATCTGCGC

* *****

* *****
* *****
* *****
* *****
* *****

EP 1401
G CGCTGCCTCC
09/471,669 1401
G CGCTGCCTCC

CCTCTTCATG CTGCCACTCT GCCTCATGGT GTGTCAGTG
CCTCTTCATG CTGCCACTCT GCCTCATGGT GTGTCAGTG

* *****
* *****
* *****
* *****
* *****

EP 1451
T CTCCCTGCTG
09/471,669 1451
T CTCCCTGCTG

GCTGCCTGCG CCAGCAGCAT GATGACTTTG CTGATGACA
GCTGCCTGCG CCAGCAGCAT GATGACTTTG CTGATGACA

* *****
* *****
* *****
* *****

EP 1501 AAGtggaggag gcccattggga gaaagataga gattccct
g ggaccacacc 09/471,669 1501 AAG-----

EP 1551 tccgtggttc actttggta caagtaggag acacagatg
g cacctgtggc 09/471,669 1504 -----

EP 1601 cagagcacct caggaccctc cccacccacc aaatgcctc
t gccttgatgg 09/471,669 1504 -----

EP 1651 agaaggaaaa ggctggcaag gtgggttcca gggactgta
c ctgttaggaaa
09/471,669 1504 -----

EP 1701 cagaaaagag aagaaagaag cactctgctg gogggata
c tcttggtcac
09/471,669 1504 -----

EP 1751 ctcaaattta agtcgggaaa ttctgctgct tgaaacttc
a gccctgaacc

09/471,669 1504

EP 1801 tttgtccacc attcctttaa attctccaac ccaaagtat
t cttctttct
09/471,669 1504

EP 1851 tagtttcaga agtactggca tcacacgcag gttaccttg
g cgtgtgtccc
09/471,669 1504

EP 1901 tgtggtaccc gggcagagaa gagaccaagc ttgtttccc
t gctggccaaa
09/471,669 1504

EP 1951 gtcagtagga gaggatgcac agtttgctat ttgctttag
a gacagggact
09/471,669 1504

EP 2001 gtataaaacaa gcctaacatt ggtgcaaaga ttgcctctt
g aattaaaaaa
09/471,669 1504

EP 2051 aaaaactaga ttgactattt atacaaatgg gggcggctg
g aaagaggaga
09/471,669 1504 -----

EP 2101 aggagagggta tacaaggac aggaaatagt gggatcaa
g ctaggaaagg
09/471,669 1504 -----

EP 2151 cagaaacaca accactcacc agtcctagtt ttagaccc
a tctccaagat 09/471,669 1504 - - - - -

EP 2201 agcatcccat ctcagaagat gggtgttgtt ttcaatgtt
t tctttctgt 09/471,669 1504 - - - - -

EP 2251 ggttgcagcc tgacaaaaag tgagatggga agggcttat
c tagccaaaga 09/471,669 1504 - - - - -

EP 2301
a agttccactt
09/471,669 1504

gctctttttt agctctctta aatgaagtgc ccactaagg

EP 2351
c tatctggAAC
09/471,669 1504

gaacacatgg aatttctgcc atattaattt ccattgtct

EP 2401 cacccttaa tctctacata tgatttaggc cagcacttg
a aaatattcct
09/471,669 1504 -----

EP 2451 aaccnnaatt tgncttgggg gctttgcngn ccaggtgct
a aaagggnntg
09/471,669 1504 -----

EP 2501 ggttaggngnc cncttnatn tnatncctna aaaggttan
n g
09/471,669 1504 -----

Alignment (FASTA format) :

=====

>EP

```
ATGGCCCCAAGCCCTGCCCTGGCTCCTGCTGTGGATGGGCGCGGGAGTGCT
GCCTGCCACGGCACCCAGCACGGCATCCGGCTGCCCTGCCAGCGGCC
TGGGGGGCGCCCCCTGGGGCTGCCGCTGCCCGGGAGACCGACGAAGAG
CCCAGGGAGGCCGCCGGAGGGCAGCTTGTGGAGATGGTGGACAACCT
GAGGGGCAAGTCGGGGCAGGGCTACTACGTGGAGATGACCGTGGCAGCC
CCCCGCAGACGCTAACATCCTGGTGGATACAGGCAGCAGTAACTTGCA
GTGGGTGCTGCCCTTACCCCTGCATCGCTACTACCAGAGGCAGCT
GTCCAGCACATAACCGGGACCTCCGGAAGGGTGTGTATGaGCCCTACACCC
AGGGCAAGTGGGAAGGGAGCTGGCACCGACCTGGTAAGCATCCCCCAT
GGCCCCAACGTCACTGTGCGTGCCAAACATTGCTGCCATCACTGAATCAGA
CAAGTTCTTCATCAACGGCTCCAACCTGGGAAGGCATCCTGGGCTGGCCT
ATGCTGAGATTGCCAGGGCTGACGACTCCCTGGAGCCTTCTTGACTCT
CTGGTAAAGCAGACCCACGTTCCAAACCTCTTCTCCCTGCAGCTTGTGG
TGCTGGCTTCCCCCTCAACCAGTCTGAAGTGCTGGCCTCTGTCGGAGGG
GCATGATCATTGGAGGTATGACCACTCGCTGTACACAGGCAGTCCTGG
TATACACCCATCCGGCGGGAGTGGTATTATGAGGTGATCATTGTGGGGT
GGAGATCAATGGACAGGATCTGAAAATGGACTGCAAGGAGTACAACATG
ACAAGAGCATTGTGGACAGTGGCACCAACCTCGTTGCCAAGAAA
GTGTTGAAGCTGCAGTCAAATCCATCAAGGCAGCCTCCACGGAGAA
GTTCCCTGATGGTTCTGGCTAGGAGAGCAGCTGGTGTGCTGGCAAGCAG
GCACCACCCCTGGAACATTTCAGTCATCTCACTCTAACCTAATGGGT
GAGGTTACCAACCAGTCCTCCGCATCACCATCCTCCGCAGCAATACCT
GCGGCCAGTGGAAAGATGTGGCACGTCCCAGACGACTGTTACAAGTTG
CCATCTCACAGTCATCCACGGGCACGTTATGGGAGCTGTTATCATGGAG
GGCTCTACGTTGTCTTGATCGGGCCGAAAACGAATTGGCTTGCTGT
CAGCGCTTGCACATGTGCACGATGAGTTCAAGGACGGCAGCGGTGGAAAGGCC
CTTTGTCACTTGGACATGGAAGACTGTGGCTACAAACATTCCACAGACA
GATGAGTCACCCCTCATGACCATAGCCTATGTCATGGCTGCCATCTGCGC
CCTCTCATGCTGCCACTCTGCCTCATGGTGTGTCAGTGGCGCTGCCTCC
GCTGCCTGCCAGCAGCATGACTTGTGATGACATCTCCCTGCTG
```

AAGtggaggaggccatggagaaagatagagattcccctggaccacacc
 tcgtggttcactttgtcacaagttaggagacacagatggcacatgtggc
 cagagcacctcaggaccctccccacccaccaaatgcctctgccttgatgg
 agaaggaaaaggctggcaagggtgggttcaggactgtacctgttaggaaa
 cagaaaagagaagaaagaaggactctgtggcggaataactcttggtac
 ctcaaatttaagtgggaaattctgtgtgtgaaacttcagccctgaacc
 ttgtccaccattttaaattctccaacccaaagtattcttctttct
 tagttcagaagtactggcatcacacgcaggttaccttggcgtgtccc
 tgggtacccggcagagaagagaccaagcttgcatttgcctgtggccaaa
 gtcagttaggagaggatgcacagttgttatggctttagagacagggact
 gtataaacacaaggctaacattggtgcaaagattgcctttgaattaaaaaa
 aaaaactagattgactattatacaaattggggcggctggaaagaggaga
 aggagaggagttacaaagacagggaatagtggatcaaagcttaggaaagg
 cagaaacacaaccactcaccagtcttagtttagacctcatctccaagat
 agoatcccatttcagaagatgggtttcaatgtttcttctgt
 ggtgcagcctgacccaaagtggatgggagggcttatctagccaaaga
 gcttttttagctctttaatgaagtgtggccactaaggaagttccactt
 gaacacatggaaatttctccatattaattccatgtcttatctggAAC
 cacccttaatctctacatatgatttaggtccagcacttggaaaatattcct
 aaccnnaatttgncnttggggcttgcnngccagtgctaaaaggnttg
 gtaggnngnccnctntatntnatnctnaaaaggttanng

>09/471,669

ATGGCCCAAGCCCTGCCCTGGCTCCTGCTGTGGATGGCGCGGGAGTGCT
 GCCTGCCCAACGGCACCCAGCACGGCATCCGGCTGCCCTGCGCAGCGGC
 TGGGGGGCGCCCCCTGGGCTGGGCTGCCCGGGAGACGACGAAGAG
 CCCGAGGAGCCCAGGGAGGGCAGCTTGTGGAGATGGTGGACAAACCT
 GAGGGGCAAGTCGGGGCAGGGTACTACGTGGAGATGACCGTGGCAGCC
 CCCCAGACGCTAACATCCTGGGATACAGGCAGCAGTAACTTGCA
 GTGGGTGCTGCCCTTACCCCTTCCTGCATCGCTACTACCAGAGGCAGCT
 GTCCAGCACATAACGGGACCTCCGGAAAGGGTGTATGTGCCCTACACCC
 AGGGCAAGTGGGAAGGGAGCTGGGACCGACCTGGTAAGCATCCCCCAT
 GGCCCCAACGTCACTGTGCGTGCACATTGCTGCCATCACTGAATCAGA
 CAAGTTCTTCATCAACGGCTCCAACCTGGAAAGGCATCCTGGGCTGGCCT
 ATGCTGAGATTGCCAGGCCTGACGACTCCCTGGAGCCTTCTTGACTCT
 CTGGTAAAGCAGACCCACGTTCCAACCTCTTCCTGCAGCTTGTGG
 TGCTGGCTTCCCCCTCAACCAGTCTGAAGTGCTGGCCTCTGGAGGGGA
 GCATGATCATTGGAGGTATCGACCACTCGCTGTACACAGGCAGTCTCTGG
 TATACACCCATCCGGGGAGTGGTATTATGAGGTGATCATTGTGCGGGT
 GGAGATCAATGGACAGGATCTGAAAATGGACTGCAAGGAGTACAACATATG
 ACAAGAGCATGTGGACAGTGGCACCAACCTCGTTGCCCAAGAAA
 GTGTTGAAGCTGCAGTCAAATCCATCAAGGCAGCCTCCTCCACGGAGAA
 GTTCCCTGATGGTTCTGGCTAGGAGAGCAGCTGGTGTGCTGGCAAGCAG
 GCACCAACCCCTGGAACATTTCCCAGTCATCTCACTTACCTAATGGGT
 GAGGTACCAACCAGTCCTCCGCATCACCACCTTCCGAGCAATACCT
 CGGGCAGTGGAAAGATGTGGCCACGTCCAAAGACGACTGTTACAAGTTG

CCATCTCACAGTCATCCACGGGCACTGTTATGGGAGCTGTTATCATGGAG
GGCTTCTACGTTGTCTTGATCGGGCCCGAAAACGAATTGGCTTGTCTGT
CAGCGCTTGCATGTGCACGATGAGTTCAAGGACGGCAGCGGTGGAAGGCC
CTTTGTCAACCTTGGACATGGAAGACTGTGGCTACAAACATTCCACAGACA
GATGAGTCAACCCTCATGACCATAGCCTATGTCATGGCTGCCATCTGC
CCTCTTCATGCTGCCACTCTGCCTCATGGTGTGTCAGTGGCGCTGCCTCC
GCTGCCTGCGCCAGCAGCATGATGACTTGTGATGACATCTCCCTGCTG
AAG-----